

Parametrizace ekvalizace elektronegativity metodou diferenciální evoluce

Jana Pazúriková¹, Tomáš Raček¹, Stanislav Geidl², Radka Svobodová Vařeková², Aleš Křenek³

PARCIÁLNÍ NÁBOJE

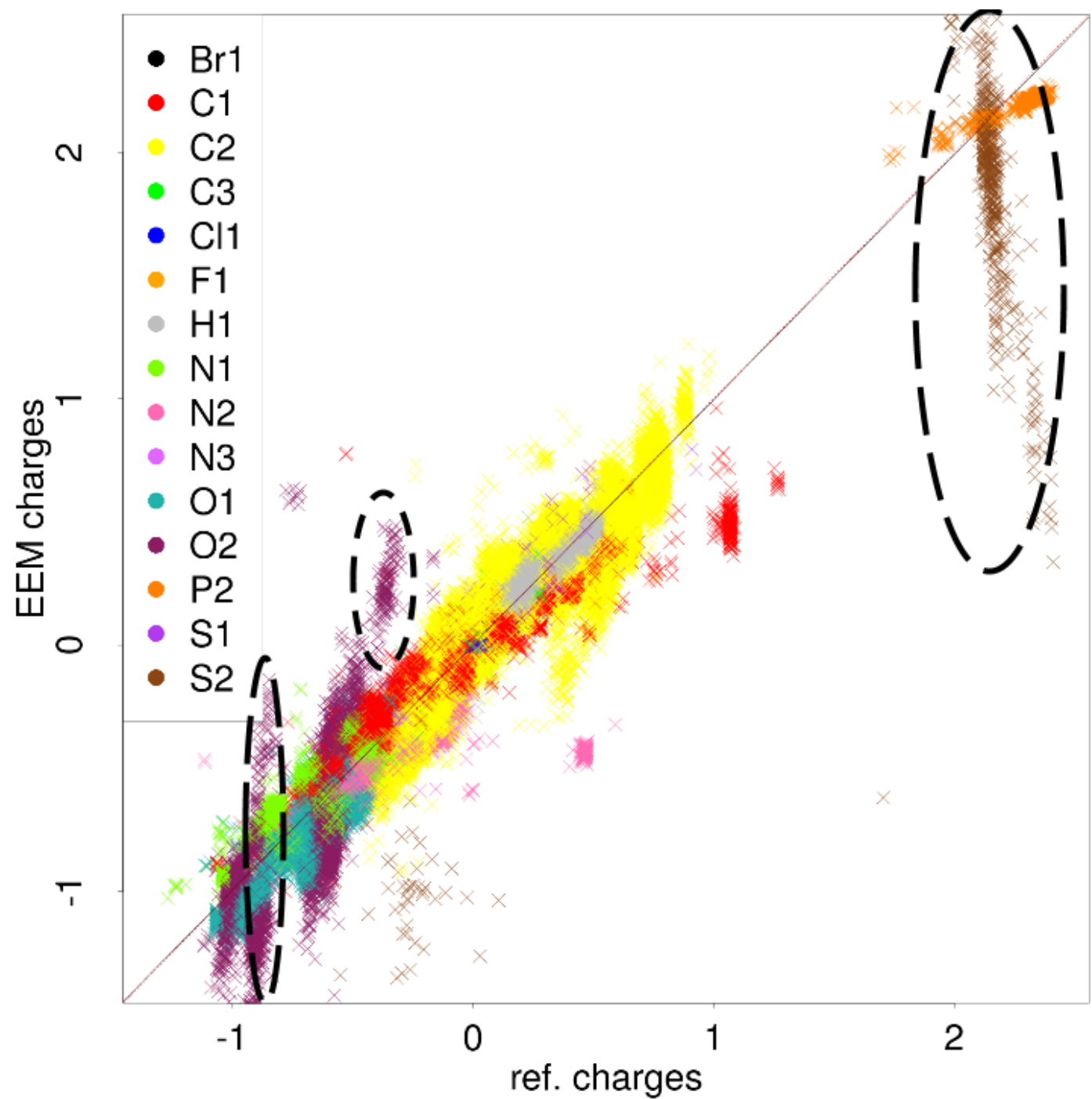
Parciální náboj vyjadřuje jedním reálným číslem rozložení elektronové hustoty v okolí daného atomu v molekule. Elektronová hustota se mění v závislosti mj. na elektronegativitě. Tyto zjednodušující teoretické koncepty se používají např. ve výpočetních simulacích a k charakterizaci vlastností molekul. Základní metody výpočtu vycházejí z kvantové mechaniky, jsou přesné, ale výpočetně náročné. Empirické metody využívají zjednodušené fyzikálně-chemické principy a empiricky zjištěné konstanty: parametry. Jsou výrazně rychlejší a mohou dosáhnout srovnatelné přesnosti.

METODA EKVALIZACE ELEKTRONEGATIVITY

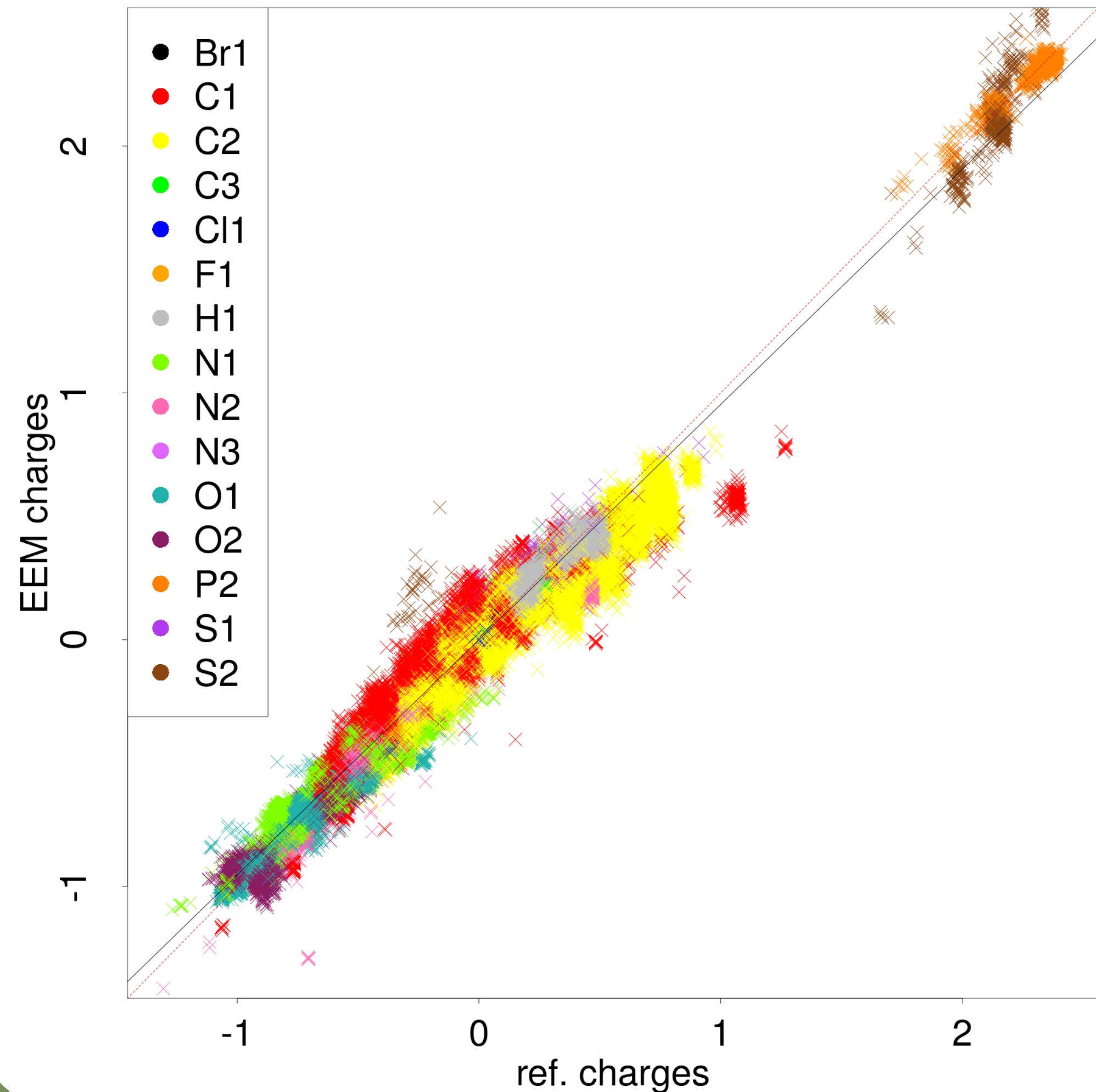
Ekvalizace elektronegativity (EEM) je jedna z empirických metod, založená na systému rovnic

$$\bar{\chi} = \chi_i = A_i + B_i q_i + \kappa \sum_{i \neq j} \frac{q_j}{r_{i,j}} \quad Q = \sum q_i \quad (1)$$

kde χ_i je elektronegativita a q_i náboj atomu i , Q celkový náboj, $r_{i,j}$ vzdálenost atomů i a j , A_i, B_i, κ parametry. Většina implementací postupně fixuje κ a počítá parametry A_i, B_i lineární regresí (LR), s cílem co nejlepší shody s náboji z tréninkové sady (přesné QM). Vypočtené parametry se použijí na výpočet nábojů v dalších molekulách. Častým problémem je ale špatná shoda vypočtených nábojů pro některé typy atomů.



PARAMETRIZACE EEM DIFERENCIÁLNÍ EVOLUCÍ



Místo lineární regrese na výpočet parametrů EEM použijeme diferenciální evoluci (DE) v kombinaci s lokální minimalizací. Algoritmus náhodně kombinuje vektory parametrů (κ, A, B pro všechny atomové typy) z předem vytvořené populace, z nového vektoru vypočte náboje podle rovnic (1) na molekulách tréninkové sady, a porovná je s přesnými náboji pomocí hodnotící funkce (odchylky nábojů, celkové nebo dílčí korelace apod.). Takto pseudonáhodně prohledáme prostor všech parametrů.

Podstatné výhody

- optimalizuje rovnocenně κ, A_i, B_i
- lépe se vyrovnává s heterogenitou velikosti a typu molekul
- všechny typy atomů vykazují dobrou shodu s QM
- lze najít více rovnocenných řešení a mezi nimi vybírat podle dalších kritérií
- lze dobře regulovat intervaly pro A a B
- je obecně rychlejší než lineární regrese

Lokální minimalizace (používáme algoritmus NEWUOA) je aplikována při tvorbě počáteční populace algoritmu DE a k finálnímu doladění výsledku. Tento postup výrazně zlepšuje kvalitu výstupu, resp. rychlost konvergence metody, aniž by narušil její globální vlastnosti.

VÝSLEDKY

Algoritmy LR a DE jsou srovnatelné v hlavním výsledku – korelaci spočtených nábojů vůči QM. DE ale dosahuje podstatně lepších dílčích výsledků. I problematické typy atomů si drží nízkou průměrnou i maximální odchylku náboje, a dílčí korelace pro daný typ je zpravidla lepší.

	DE R2	DE RMSD	DE Δmax	LR R2	LR RMSD	LR Δmax
H1	0.858	0.0328	0.1684	0.8814	0.0348	0.3076
C1	0.9076	0.0925	0.6347	0.9269	0.0905	1.3035
C2	0.9321	0.0945	0.4552	0.8936	0.1061	0.9727
C3	0.8495	0.063	0.2227	0.8314	0.0671	0.2729
N1	0.8055	0.0747	0.2997	0.7201	0.0901	0.5395
N2	0.8851	0.0697	0.5899	0.5038	0.1803	1.5469
N3	0.4183	0.0151	0.0818	0.3753	0.0156	0.0864
O1	0.8897	0.0463	0.3674	0.7687	0.1055	0.4279
O2	0.9162	0.0506	0.2089	0.5956	0.2145	4.1639
F1	0.3787	0.034	0.1273	0.1209	0.0246	0.1655
P2	0.9158	0.0353	0.1742	0.8948	0.1106	0.4213
S1	0.7455	0.0787	0.2781	0.7383	0.08	0.2633
S2	0.9399	0.1595	0.698	0.4615	0.7908	7.6148
Br1	0.4222	0.0268	0.0819	0.4506	0.0358	0.1143
Cl1	0.3526	0.0226	0.1332	0.3458	0.0287	0.1212

Legend: RMSD < 0.05 < 0.1 < 0.15 < 0.2 < 0.3 < 0.4 >= 0.4

R2 – Pearsonův korelační koeficient, RMSD – průměrná odchylka spočteného náboje, Δmax – maximální odchylka náboje

ALGORITMUS

```
najdi_parametry_DE() {
1: populace ← vytvor_populaci()
2: ∀ x ∈ populace: R(x) ← ohodnot(x)
3: ∀ x ∈ populace && R(x) > 0.3: newuoa(x)
4: nejlepsi ← najdi_nejlepsi(populace)
5: loop
6:   novy ← kombinuj(vyber_nahodny(populace),
                     vybr_nahodny(populace))
7:   R(novy) ← ohodnot(novy)
8:   if R(novy) > R(nejlepsi) then
9:     nejlepsi ← novy
10:  end if
11: end loop
12: newuoa(nejlepsi)
}

ohodnot(x) {
1: ∀ molekula ∈ treningova_sada: eem_naboje ← eem(x)
2: R ← porovnej(eem_naboje, qm_naboje)
}
```

[EEM] W. J. Mortier, S. K. Ghosh, and S. Shankar, Electronegativity-equalization method for the calculation of atomic charges in molecules, Journal of the American Chemical Society, vol. 108, no. 15, pp. 4315–4320, 1986.
[NEWUOA] Powell, M. J. D. (2004). "Least Frobenius norm updating of quadratic models that satisfy interpolation conditions". Mathematical Programming (Springer) 100: 183–215.
[DE] R. Storm, K. Price, Differential Evolution: A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, Journal of Global Optimization, Volume 11, Issue 4, pp. 341-359, 1997

¹Fakulta informatiky, MU ²CEITEC, MU ³Ústav výpočetní techniky, MU

Tato práce vznikla za podpory projektem CERIT Scientific Cloud LM2015085, s využitím výpočetních zdrojů poskytnutých projektem LM2015042 E-infrastruktura CESNET.